DOCUMENT RESUME

ED 040 952 SP 003 999

AUTHOR	Luft, Max; Bemis, Katherine A.
TITLE	Video Tape Techniques for Establishing Inter-Rater Reliability.
INSTITUTION	Southwestern Cooperative Educational Lab., Albuquerque, N. Mex.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
PUB DATE	Mar 70
CONTRACT	OEC-4-7-062827-3078
NOTE	7p.; Paper presented at annual meeting, AERA, Minneapolis, 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.45
DESCRIPTORS	*Classroom Observation Techniques, Interaction Process Analysis, *Reliability, Video Tape Recordings

ABSTRACT

The object of this study was to validate a technique for establishing inter-rater reliability on the Southwestern Cooperative Interaction Observation Schedule (SCIOS), where it was impractical to bring the observers to a common site. Reliability was originally obtained when eight observers met together. Observers were divided into four pairs. A video tape of a typical classroom scene was transported to each of the pairs in four cities. All observers viewed the same tape within a one-week period. Correlations of each observer with all others were averaged. This average correlation was compared with correlations of observers normally working together. The mean of all the correlations was .457. The corresponding mean for correlations of paired observers was .904. This technique proved to be superior financially and statistically in discriminating need for further training of observers to obtain inter-rater reliability, as compared to using only correlations of paired observers or bringing all observers to a common site. (Author)

MAY 27 1970

ED040952

VIDEO TAPE TECHNIQUES FOR ESTABLISHING INTER-RATER RELIABILITY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Dr. Max Luft and Dr. Katherine A. Bemis
Southwestern Cooperative Educational Laboratory, Inc.
Albuquerque, New Mexico

This publication is issued pursuant to terms of Contract No. OEC-4-7-062827-3078 with the Bureau of Research, Office of Education, U. S. Depoartment of Health, Education and Welfare.

A paper presented at the Annual Convention of the American Educational Research Association, Minneapolis, Minnesota, March 2-6, 1970.

SP003999

VIDEO TAPE TECHNIQUES FOR ESTABLISHING INTER-RATER RELIABILITY

Dr. Max Luft and Dr. Katherine A. Bemis
Southwestern Cooperative Educational Laboratory, Inc.
Albuquerque, New Mexico

Classroom observations mean many things to many people. Administrators observe in the classroom in order to evaluate the teachers. Student teachers observe in the classroom in order to learn new teaching techniques. At Southwestern Cooperative Educational Laboratory (SWCEL) our classroom observations have another purpose. We are attempting to link the interaction between students and teachers with pupil gain on cognitive materials.

The observational instrument, the Southwestern Cooperative Interaction Observation Schedule (SCIOS), measures verbal and nonverbal interaction between the students and teachers as it occurs during a sixteen minute interval. Very loosely, it may be stated that the schedule is an attempt to measure affective attitudes of the teacher toward the students and their reactions to the teacher. The achievement tests used as cognitive measures to establish relationships with the SCIOS were the subtests of the California Achievement Test.

It is one of the purposes of SWCEL to facilitate application of theory to the classroom. A key administrator of one of the regional laboratories has said that, "An educational laboratory uses known research to bridge the gap between theory and practice--and while some basic research may be conducted, most of the efforts are aimed at applied research and directed towards the marriage of content and implementation of the procedures." (Olivero, 1968). It was the purpose

of this study to find existing relationships between teacher attitude and cognitive behavior, that teacher attitude might be varied in an in-service training summer institute in such a manner that it would improve cognitive achievement of students on a standardized test.

The problem arose in trying to implement the observation schedule in four southwestern cities (Phoenix, Arizona; Bernalillo, New Mexico; Odessa, Texas; Tulsa, Oklahoma) where the problem of bringing the observers together to view a common classroom situation was impractical because of cost and time factors. There were two observers working on a part time basis at each site.

It was during the 1968-69 school year that SWCEL worked with a pair of observers in each of these cities who observed in a total of 126 classrooms every other week from mid October through the end of April. The specific problem of this paper deals with the establishment of inter-rater reliability between these eight observers in the four cities during the school year.

PROCEDURE

The eight observers were selected by SWCEL from employment applications to SWCEL, from employment applications to local school districts, and from recommendations by local university professors. Qualifications for the observers were that they were not currently employed by the participating school district, they could spend at least fifteen to twenty hours a week working for the Laboratory, and had satisfactory elementary school teaching experience.

The eight observers were brought to SWCEL, in Albuquerque, during the first week of September. They were given an introduction into the

techniques of making a classroom observation, and they were familiarized with the observation schedule itself. They viewed a series of video tapes of non-experimental classrooms to become more familiar with the observation schedule. They were then sent to observe teachers in the local schools who had participated with the Laboratory previously in classroom observations, and possessed little anxiety about being observed. During a series of six trial periods, the observers were paired randomly with each other.

The observation procedure was to visit the classroom, sit in approximately the same area, and using a stop watch, observe the students for the sixteen minute required interval. Following the observation, a discussion was held, away from the classroom, about differences that existed in scoring of the observation schedule between each of the paired observers. With increased observations, discrepancies became fewer. Observers were paired with the observers who would be working with them in their local city. At the end of the training period in Albuquerque, an average was taken of the final trial of four-paired observations. The average of these four correlations was .834.

It was anticipated that with more observational experience in pairs, individual correlations would tend to increase. The question that arose is; Would the four pairs of observers who, although, they were agreeing more with each other still have had as high a correlation with other observer pairs? It was impractical to bring all eight of the observers periodically to Albuquerque. However, the Laboratory personnel did visit each of the four sites bi-monthly to insure the quality of the program going on in the classroom. It had been decided that it would be more feasible, economically, to transport video tapes

to the four sites and have the observers scan an identical portion of video tape for the sixteen minute period, and then correlate their results in Albuquerque. Results would be correlated between pairs and also in an eight by eight correlation matrix to find out exactly which people were not viewing the tape in the same manner.

Observers were instructed to make contact with the local school system as soon as they arrived from their training in Albuquerque and request permission to continue observing the classrooms as often as possible during the first month. The purpose of the increased observations during the first month was to gain greater inter-rater reliability between the two observers in each of the four sites. Care was taken, however, that teachers participating in the study were observed only every other week, with other observations being made in classes not related to the study.

It was intended to discard the first three observations because it was felt that the teachers would probably be anxious on the first few observations and that the activity would not represent a normal situation. It was assumed that as the teachers became accustomed to the observer's presence in the classroom, they would become "their natural selves." It was decided to send a video tape to the field one month after the observers had left the Laboratory. All teams had, by this time, made at least twenty-five paired observations. The video tape was then viewed by each observer in the field. The tape was marked as to when the observation should begin and end. The video tape was made on half-inch Sony equipment, and viewed by all observers during the same week.

The observation schedules were tallied and the data were correlated. The four observer pairs had correlations with each other of .935, .893, .894, and .897 on the thirty-eight items in the schedule. The average correlation for all pairs was then .904. This indicated that the observers were now more highly correlated than what they were seeing, but the question still remained as to whether they were really establishing higher reliability between pairs of observers. The lowest correlation on the eight by eight matrix was .071 with the highest correlation between any two nonpaired observers being .574. The average of the correlations in the eight by eight matrix was .457.

The difference between .457 as the average of all correlations, and .904 as the average of paired observations was greater than had been expected. It was decided to revisit each of the sites and to go over the correlations with each of the observers indicating the differences between the observation schedules. At the same time, the designer of the observation schedule met with the observers to examine discrepancies with them. One observer was also isolated by this technique by having exceptionally low correlations with six of the observers. She was designated as one who needed additional help and time was spent in instructing her in filling out the observation schedule. An interesting side light to note is that her paired observer had extremely high correlations with the other six observers with whom she did not usually participate.

RESULTS

The results of this technique of establishing inter-rater reliability may be assessed both financially and statistically in discriminating the need for further training of observers in obtaining

inter-rater reliability. As the case point, one of a pair of observers had a very low reliability with the other six, while her mate had very high reliability. This indicates the impracticality of looking only at paired observations.

The advantages of training classroom observers using video taped classroom scenes are financially and empirically unlimited. Having a laboratory staff member visit each of the four sites to re-establish inter-rater reliability eliminates the financial burdens of attempting to bring eight observers from four states to a common site. In using the video taped classroom scenes, we were also certain that all of the observers were watching exactly the same classroom phenomena at the same time. Even if we had been able to bring all of the observers to the same site, we would have had to have many observations conducted in a variety of classrooms in order to get the number of combinations made possible by using the video taped classroom scenes.

After inter-rater reliability was established among all eight observers, a significant relationship was found to exist between certain SCIOS behaviors and student achievement scores. The implications of this finding for educators and students are infinite.